

This is a pre-print version of an article in *Studies in Second Language Acquisition* 27 (2005)
Please consult the published version if citing.

Language transfer and discourse universals in Indian English article use

Devyani Sharma

ABSTRACT

Stable non-native varieties of English acquired and used in the absence of native English input can diverge systematically from native varieties over time (Kachru 1983; Platt, Weber, and Ho 1984; Cheshire 1991). Focusing on Indian English article use, this study asks: if divergence is indeed occurring, do new features derive primarily from L1 transfer or from universal principles? Natural conversational speech is assessed in relation to four hypotheses relating to L1 transfer and language universals, and a multivariate regression analysis evaluates the relative strength of each factor. The new article system is not found to be identical to the L1 article system. Although L1 transfer appears to be operative when an overt form (the specific indefinite article) exists in the L1, when a gap occurs in the L1 (no definite article) speakers do not completely omit the definite article in their L2 English. Using Prince's (1981) Taxonomy of Assumed Familiarity, I show that the absence of an L1 model for definite articles permits the intervention of universally available discourse knowledge, such that speakers apply an economical, disambiguating principle to the use of overt articles, reserving them mainly for new (less given or inferable) information and omitting them in more redundant contexts.

INTRODUCTION

Second languages spoken in stable bilingual settings, for instance English as spoken in India or Singapore, can become indigenized over time due to the historical removal of the native variety and broad societal transmission of second language traits through institutional as well as informal channels. The forces that influence the direction of change in such cases are similar to forces active in classroom and individual second language acquisition (SLA), but their ultimate impact may be much more lasting due to the increasing function of the second language variety as a dialect in its own right. Native varieties of English are no longer the target language for many such speakers, and instead a standard version of the indigenized variety performs this function (Kachru, 1983; Platt, Weber and Ho, 1984; Sridhar, 1985; Lowenberg, 1986; Sahgal, 1991). As Kandiah (1991), Bamgbose (1998), and others have noted, if stable non-native varieties of English indeed constitute a special case of second language acquisition, then the distinctiveness of their grammatical systems must be shown through systematic analysis.

Explanations for SLA patterns in L2 English article use have included first language transfer (Platt, Weber and Ho, 1984; Jarvis, 2002), universal UG distinctions (Bickerton, 1981; Ionin, Ko, and Wexler, 2003), topic continuity (Huebner, 1983; Chaudron and Parker, 1990; Jarvis, 2002), and conversational genre (Tarone and Parrish, 1988). Language transfer in particular has also frequently been invoked to account for article use in New Englishes, although these claims have not always been subjected to close scrutiny. The present study evaluates the relative importance of language transfer

and linguistic universals in the development of new principles for article use in the natural speech of Indian English speakers. Within the domain of universals, a particular interest of this study lies in the reliance by L2 speakers on knowledge of the discourse status of entities in the absence of a clear native target model of semantic and pragmatic principles of article use. The findings strongly suggest that while language transfer does play a role, particularly when the first language has an overt form designated for a particular function (as opposed to lacking a form that exists in the L2), speakers appear to exploit universally available discourse knowledge to a significant extent as well in order to create a new functional basis for the use of articles.

After a discussion of data collection, I first construct an implicational scaling of speakers according to a number of syntactic variables, in order to present an introductory sketch of their grammars. As will be evident from this initial scaling, articles represent a notably more divergent part of Indian English syntax as compared to many other syntactic features. Next, I propose four competing hypotheses that might account for the new patterns of article use: the first three hypotheses have been examined in other studies, while the fourth introduces a new potential factor. I follow common sociolinguistic practice in coding independent factors for each token, and using this coding, I first evaluate each hypothesis and then compare the relative importance of all factors using a multivariate statistical regression analysis.

DATA COLLECTION

Unlike many studies of indigenized non-native English, which for various reasons have been limited to relatively proficient speakers, the present study encompasses a wide range of proficiency in English in order to reflect more accurately the complex nature of bilingual continua. The data were collected among first generation adult Indian immigrants in California. All twelve individuals acquired English to varying degrees in India, emigrated as adults to the United States, and for the most part maintained their multilingual repertoires in the United States. At the time of the study, most of these subjects were employed in small commercial shops and two were working in software companies, all within the San Francisco Bay Area (specifically Berkeley and Sunnyvale). Restricting the study to individuals who were generally part of Indian immigrant social networks and were not involved in academic study to some extent minimized the factor of formal instruction in American English. Their increased informal contact with a native variety of English is of course an important new development in their linguistic environment: they show socially-motivated variation in their patterns of adoption of American phonological features in particular (see Sharma, 2005). However, for the present discussion I focus on article use and do not specifically address the role of contact with American English.

Table 1 lists background information for all speakers, ordering them according to the results presented later in Table 2. The twelve speakers do not comprise an exact cross-classification of every external variable; however, since the goal was to broadly characterize features of the Indian English bilingual continuum, this shortcoming is balanced by the inclusion of a relatively diverse range of English proficiency levels in the group.

Table 1. Participants' social characteristics

Characteristics	Participants											
	KD	SK	CK	RS	RR	KP	KK	GV	RT	KB	SS	NT
Education	0	0	0	1	1	1	2	2	2	2	2	2
Functional role	0	0	0	1	1	2	2	1	2	3	3	3
Years in the United States	2	18	17	2	17	25	39	0.5	0.5	40	0.7	2
Age	34	38	67	26	48	54	62	35	29	67	23	24
Sex	M	F	M	M	M	F	M	M	M	M	M	M
L1 ^a	Guj	Pun	Guj	Tam	Guj	Guj	Pun	Kan	Hin	Hin	Hin	Hin

Note. Education: 0 = no formal English; 1 = English in higher education; 2 = primarily English education; functional role: 0 = minimally at work; 1 = regularly at work; 2 = work and some friendships/younger relatives; 3 = work, friendships, home; L1: Guj = Gujarati; Pun = Punjabi; Tam = Tamil; Kan = Kannada; Hin = Hindi/Urdu.

^aMost individuals speak Hindi in addition to their L1.

As a central interest of the present study lies in characterizing the natural development and use of new non-native dialect features, the data were collected through relatively naturalistic sociolinguistic interviews rather than through elicitation tasks. The interviews, ranging in duration from one to two hours, were aimed at eliciting naturalistic speech data, personal demographic information, personal experiences and narratives, and information about the speakers' attitudes towards language use, dialects and cultural contact. Particularly for the question of how article use may vary according to discourse structure, the relatively informal nature of speakers' responses was advantageous. This stands in contrast to many studies of article use that have relied on elicitation tasks or grammaticality judgements (Agnihotri, Khanna and Mukherjee, 1994; Goto Butler, 2002; Ionin, Ko, and Wexler, 2003; Y. Kachru, 2003). Experimental tasks of this type are more controllable, and thus can extract data for a complete range of contrasting article contexts; however, they almost invariably place speakers in highly formal and standard linguistic interactions, thus potentially eliciting maximally standard usage and possibly even distortions of natural usage due to the speaker's heightened linguistic self-consciousness in the test situation. While these types of controlled elicitation tests may be appropriate for pedagogically motivated inquiry, they can have a skewing impact on any attempt to investigate speakers' natural, unmonitored use of the language.

IMPLICATIONAL SCALING OF GRAMMATICAL FEATURES

B. Kachru (1965, p. 393-6) developed the concept of a 'cline of bilingualism' in order to introduce multidimensionality to the study of indigenized varieties of English, in the form of variability at different levels such as acquisition, function, and context of situation. While frequent reference has been made to this cline, few studies have sought direct empirical evidence in the natural speech of bilingual Indian English speakers. As implicational scaling was introduced into the study of creoles and of SLA in order to identify different dimensions of variation, it is particularly well-suited to a quantitative implementation of Kachru's notion (Cf. Ho and Platt, 1993; Agnihotri, Khanna and Mukherjee, 1994).

In an implicational relationship, the presence of a feature \underline{x} implies the presence of \underline{y} , but not vice versa. Thus, one may have a grammar that includes only the feature \underline{y} , or a grammar that includes both features \underline{x} and \underline{y} , but a grammar that exclusively includes the feature \underline{x} is not predicted. In tabular form, this relationship translates into the claim that a value in one column will have higher values above and to its left, and the reverse will be true for lower values. Although implicational scaling is the subject of some debate in both creole studies (DeCamp, 1971; Bickerton, 1971; see Rickford, 2002 for a summary) and in SLA (Andersen, 1978; Huebner, 1983; Hudson, 1993; Pienemann, Johnston, and Meisel, 1993; Bayley, 1999), I adopt it here merely as a preliminary organising tool that highlights, rather than explains, individual differences across a speech continuum.

As a first step of this work is to first identify any evidence of systematic divergence in stable non-native usage, the initial set of grammatical variables includes general 'interlanguage' features as well as features that have been specifically noted for Indian English. Variable past tense marking, copula use, and subject-verb agreement are common features of SLA (Dulay and Burt, 1974; Wolfram, 1985; Bayley, 1994; R. Hawkins, 2001), as is variation in article use. The latter in particular has been noted as characteristic of Indian English as well (B. Kachru, 1983; Platt, Weber, and Ho, 1984; Williams, 1987; Agnihotri, Khanna, and Mukherjee, 1994).

Table 2 lists percentage rates of non-standardness for each of these features: past tense (e.g. I stay(ed) in San Francisco last year), copula (e.g. He (is) crazy), subject-verb agreement (e.g. Our prices is (are) cheaper), definite article (e.g. (The) driver gave me directions), non-specific indefinite article (e.g. I'm looking for (a) job), and specific indefinite article (e.g. I met (a) friend of yours). Specificity, definiteness, and details of article usage are explicated in greater detail in the section that follows.

Table 2. Percentage rates of nonstandard forms by speaker

Speaker	Scale A						Scale B					
	No past marking		No copula		Agreement mismatch		No definite article (evoked)		No indefinite article (nonspecific)		No indefinite article (specific)	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
KD	30	70	112	24	112	6	7	86	17	82	6	50
SK	66	70	63	21	102	16	4	50	19	84	12	50
CK	19	47	31	23	85	9	—	—	6	67	11	45
RS	169	28	109	16	127	3	33	55	38	58	32	31
RR	90	20	134	5	162	11	5	60	37	57	26	46
KP	78	22	106	5	130	8	9	78	19	47	13	23
KK	83	7	92	15	98	4	8	50	55	65	30	30
GV	76	3	130	2	112	0	13	62	57	51	28	21
RT	68	0	96	0	116	0	10	50	18	44	20	20
KB	86	0	93	0	122	0	5	40	28	29	53	21
SS	30	0	69	0	72	0	4	75	16	25	11	36
NT	131	0	112	0	134	0	4	0	70	9	32	0

Note. Horizontal scalability: 91.6% (columns 1–3), 86.1% (columns 4–6); vertical scalability: 86.1% (columns 1–3), 69.4% (columns 4–6). *n* values in all tables report the total number of relevant contexts in which nonstandard use of a variable might occur.

The scalability of an implicational distribution, calculated by dividing the number of correctly predicted values by the total number of values, represents the closeness of fit between the data and the predicted implicational model. The high rates of scalability in both dimensions in Table 2 demonstrate first, as many previous studies have, that second language speech is clearly structured despite considerable variation across speakers. Particularly in the horizontal dimension, of a total of 72 cells in the table, only eight values violate the predicted ordering; furthermore, these violations generally remain within the expected range and do not diverge dramatically from the values in the neighboring cells.

Perhaps the more important finding in Table 2, however, is that the data can in fact only be arranged into two separate implicational scalings, because the variables conform fundamentally to strikingly different patterns of variation.

The features in ‘Scale A’ on the left side of Table 2 — subject-verb agreement, past marking, and copula use — converge towards native English-like usage halfway down the continuum, so that the more proficient half of the group do not show any indigenized usage of these features. In keeping with many learning trajectories, the distribution across these first three columns resembles an S-curve, whereby the majority of the values (34/36) in the distribution are restricted to lower (0%-25%) and higher (65%-100%) frequency ranges, and a minority (2/36) fall in a transitory intermediate frequency range (25%-65%). It is worth noting that the actual ordering of these three interlanguage features is different to ‘universal’ orders of acquisition proposed for English (Cf. Dulay and Burt 1974), and I make no claim that the order of interlanguage features apparent here is universal.

The considerable overlap between English and the first languages of these individuals with respect to these particular features may reinforce the ultimate acquisition of the native English system. All five first languages mark past tense overtly, in some cases with more distinctions (e.g. in gender and number) than English; all five also include some system of subject-verb agreement and, again, these systems are often more elaborated than English (e.g. person, number and gender agreement and object agreement); and finally, all five languages have copular verbs, although Tamil and Kannada do permit null copula in certain clauses, such as NP-NP constructions.

The features in ‘Scale B’ on the right side of Table 2, by contrast, are more systematically divergent, persisting long after other L2 learning features have been acquired.¹ Many of these uses have come to be generalized to the extent that they appear in the speech of individuals who consider English to be

one of their native languages (KB, SS, NT). Although gradually decreasing across the continuum of speakers, these features do not follow an S-curve trajectory; instead, they exhibit a flatter and higher distribution. This could either be taken as evidence of a new non-standard article system beginning to be indigenously transmitted or as evidence of very late stage SLA; regardless of the interpretation, we can minimally conclude that the two scalings in Table 2 offer empirical evidence of two distinct patterns of variation within the grammars of these stable bilinguals.

The L1 article systems of the speakers are far more dissimilar to English than the L1 systems of tense, copulas, and agreement, and at first glance this might suggest that the greater divergence in article use derives from negative transfer from L1 systems. The remainder of this paper aims to evaluate whether this is in fact the case, and if not, what the precise nature of the developing system is.ⁱⁱ

HYPOTHESES FOR INDIAN ENGLISH ARTICLE USE

Two core dimensions of cross-linguistic variation in article systems are specificity and definiteness. Specificity signals the existence of a unique real world referent for a noun phrase, or the ‘speaker’s ability to identify the referent’ (Fodor and Sag, 1982). A non-specific NP could refer to any token of its type in the real world (*I’m looking for a book. Any book will do*), while a specific NP has a unique referent in the real world (*I’m looking for a book. I think I left it here yesterday*). Definiteness, by contrast, is primarily rooted in discourse. J. Hawkins (1978) describes the definite article as an instruction for the hearer to ‘locate’ the referent of that NP within a pragmatically defined set of objects that are part of the shared speaker-hearer knowledge. This can be referred to as the ‘givenness’ of information that the speaker can treat as ‘recoverable either anaphorically or situationally’ (Halliday 1967). A definite NP has already been referred to by a speaker and can be assumed to be known (*I asked a boy for directions. The boy told me*), while an indefinite NP has not been referred to by a speaker and cannot be assumed to be known (*I asked a boy for directions*).

L2 article use may be influenced by transfer of an L1 system of marking definiteness and specificity, or by universal preferences in the linguistic marking of these categories. The first two hypotheses below relate to language transfer, while the latter two present potential universal factors.

Hypothesis I: Transfer of L1 positional marking of discourse status

The Indo-Aryan L1s of the present speakers lack a definite article and mark definiteness via word order (Mohanani, 1994; McGregor 1995), a system exemplified in the examples in (1).

- (1) a. [kitaab] [mez par] [pari hai]
 book table on lying is
 ‘The book is lying on the table.’
- b. [mez par] [kitaab] [pari hai]
 table on book lying is
 ‘A book is lying on the table.’
- c. [mez par] [pari hai] [kitaab]
 table on lying is book
 ‘On the table is lying a book.’

The definiteness interpretation of the NP *kitaab* in (1) varies according to its position in the clause. Discourse configurability of this type frequently places information familiar to the hearer (thematic information) clause-initially and new (rhematic) information clause-finally. This cross-linguistically observed pattern has been variably described as topic/focus, topic/comment, and theme/rheme (Dik, 1978; Vallduví, 1992; Lambrecht, 1994; Birner and Ward, 1998).

One type of L1 transfer in Indian English, then, might be a reliance on clause position to mark discourse status. In a study of English acquisition by speakers of Czech and Slovak, both discourse configurational languages, Young (1996) predicts that discourse items in unexpected (noncanonical)

clausal positions will be overtly marked, and redundant marking will be avoided elsewhere. Given the similar discourse configurationality of Hindi, I adopt Young's (1996, p. 156) hypotheses here:

Hypothesis I (Transfer of L1 positional marking):

- (i) Thematic NPs in clause-initial position will favor zero articles. (unmarked)
- (ii) Rhematic NPs in clause-final position will favor zero articles. (unmarked)
- (iii) Thematic NPs in clause-final position will favor definite articles. (marked)
- (iv) Rhematic NPs in clause-initial position will favor indefinite articles. (marked)

This predicts that speakers will only use articles when an NP is not in its pragmatically-determined clause position, i.e. in situations (iii) and (iv). Old information in the expected clause-initial theme position will favor zero articles, as will new information in the clause-final rheme position.

Hypothesis II: Transfer of L1 article system

Direct transfer from a distinct L1 article system is perhaps the most frequently cited explanation of divergence in article use in non-native varieties of English. As a result of extensive areal convergence, the article systems of the five L1s in the present data are fairly similar: none of the languages has a definite article, definiteness can be marked by word order or by case-marking, and all of the languages may use the numeral *one* with specific indefinite meaning (McGregor, 1995; Bhatia, 1993; Sridhar, 1990; Cardona, 1965; Schiffman, 1999). In other words, the article systems of these languages contrast specific/non-specific but not definite/indefinite. As the simplified diagrams in Figure 1 and Figure 2 show, this system is the inverse of the article system of English, which prioritizes the definite/indefinite distinction over specific/non-specific.

DEFINITE	INDEFINITE	
the	a	SPECIFIC
the, a (generic)	a	NONSPECIFIC

Figure 1: English article system
(for singular count nouns)

DEFINITE	INDEFINITE	
—	ek (one)	SPECIFIC
— (generic)	—	NONSPECIFIC

Figure 2: Hindi article system
(for singular count nouns)

Platt, Weber and Ho (1984, p. 53-9) suggest that transfer from L1 specificity-marking is the primary factor in divergent article systems in New Englishes, citing two Indian English examples, repeated in (2), that reflect direct adaptation of L2 forms—*one* and *Ø*—to L1 functions:

- (2) a. I'm staying in one house with three other students. (specific)
- b. I want to spend some time in Ø village, definitely if I get Ø chance. (non-specific)

Two distinct predictions underlie transfer from a Hindi-like system (Fig. 2) to English (Fig. 1). First, the existence of a specific indefinite article in Hindi should lead speakers to restrict their English indefinite article use to specific NPs only. Second, speakers may avoid articles with definite and generic NPs in their English, as Hindi has no such marker:

Hypothesis II (Transfer of L1 article forms and functions):

- (i) a. [+ specific, – definite] NPs will be marked overtly with an article (possibly using the form *one* rather than *a*)
- b. [– specific, – definite] NPs will have no article
- (ii) a. [+ specific, + definite] NPs will have no article
- b. [– specific, + definite] NPs will have no article

Hypothesis III: Universal tripartite article system

A mismatch between the L1 and L2 grammars may permit the intervention of universally unmarked or more ‘natural’ grammatical systems. Bickerton (1981) offers an explicitly formulated claim regarding universal distinctions in article use. He suggests that the semantic distinction of [\pm specific reference] and the discourse distinction of [\pm hearer-known] are universal and expressed by various means cross-linguistically. He argues that whereas the main parameter of the article system of standard English is definite-indefinite, as discussed above, creole systems generate an additional specificity distinction to achieve a universally preferred tripartite system: (i) [+SR, +HK], i.e. referential definites, requiring the definite article; (ii) [+SR, –HK], i.e. referential indefinites, requiring the indefinite article; and (iii) [–SR, –HK] or [–SR, +HK], i.e. non-specifics and generics, requiring zero marking.

Bickerton’s model has been adopted and evaluated in several studies of article use in SLA and in new varieties of English. Chaudron and Parker (1990) found that Japanese speakers redistributed English forms to maintain the definiteness and topic continuity distinctions of [+SR, +HK, +topic], [+SR, +HK, –topic], and [+SR, –HK]. Huebner (1983), examining a single individual’s acquisition of the English article system, also finds some evidence of Bickerton’s categories, although they are reflected in various developmental stages the learner moves through, rather than as a static, unified system. Mesthrie (1992, p. 205) suggests that sporadic article use in South African Indian English is also reminiscent of a prototypical creole system, although the three examples he furnishes lend equal, if not more, support to a direct transfer of forms from the L1 system, as Bickerton would predict a distinct marker for each of the three types.

- (3) a. I was feeling thirsty, so I bought one soda water. (+SR, –HK; a \Rightarrow one)
 b. Ø Food is lovely. (+SR, +HK; the \Rightarrow Ø)
 c. Because if they give us Ø chance... (–SR, –HK; a \Rightarrow Ø)

The universalist or typological prediction for article use based on Bickerton’s (1981) distinctions of [\pm specific] and [\pm hearer-known] can be summarized as follows:

Hypothesis III (Universal prototypical tripartite division):

- (i) [+SR, +HK] NPs will have a definite article
- (ii) [+SR, –HK] NPs will have an indefinite article
- (iii) [–SR, –HK] and [–SR, +HK] NPs will have no article

Notice that of these predictions, (ii) and (iii) hold true for the article systems of Indian L1 languages and are therefore identical in those respects to Hypothesis II; as we will see, this renders distinguishing between transfer and universal influences potentially problematic.

Hypothesis IV: Universal discourse knowledge

A second type of universal influence on L2 article use could come from discourse pragmatics rather than semantic contrasts. The status of a noun phrase as [+hearer-known] or [–hearer-known] in the previous section is a somewhat idealized dichotomy (J. Hawkins, 1978; Prince, 1981; Lambrecht, 1999; Lyons, 1999). If we consider the ‘newness’ of an NP as scalar rather than discrete, then the relative position of an NP along this scale may affect the use of articles:

Hypothesis IV (Universal discourse knowledge):

Speakers’ use of articles will vary according to the relative ‘givenness’ or ‘newness’ of the NP in question.

Jarvis (2002, p. 388) points to the important difference between the status of an NP as the topic or comment of a clause, and its status as relatively new, known, or current in the ongoing discourse. He observes that Young (1996) defines topics as old information, thus conflating potentially distinct factors of clause-level and discourse-level information. I follow Jarvis in distinguishing between these

two types of NP status: Hypothesis I will examine the topic (theme) or comment (rheme) status of an NP in its clause, while Hypothesis IV examines the overall discourse status of the NP.

CODING METHODOLOGY

This study examines omission of the and a in contexts that would standardly require an overt article in English. Contexts in which standard English would not require an article were therefore not included in the coding, for instance, indefinite noncount nouns (e.g. I need furniture) and all plurals, as indefinite plurals do not occur with overt articles (e.g. I teach children), and generic and definite plurals have optional articles (e.g. (The) prices went up). Extraneous article insertion in contexts that standardly do not require an article (e.g. We speak the Hindi) did occur occasionally but these uses are not discussed here either. Following Huebner (1983), Tarone and Parrish (1988), and Young (1996), I excluded possessives, numerals, demonstratives, and quantifiers, mainly due to the lack of significant interaction between these determiners and the and a. Two sets of tokens were therefore coded: the indefinite article a (presence/absence) and the definite article the (presence/absence).ⁱⁱⁱ

I use the term ‘null’ to refer to absence of a or the when either article would be standardly overt. Chesterman (1991) refers to the non-overt indefinite article (e.g. with mass and plural nouns) as the zero article, and the non-overt definite article (e.g. with proper nouns) as the null article. Adopting either of these meanings would be problematic, as both refer to standard article absence in English, and the article absence of interest here is non-standard. I therefore use the term null in a different sense, simply to indicate the non-standard absence of either article.

In one set of data, the dependent variable was the absence of the indefinite article, occurring in contexts where Standard English^{iv} would require the overt indefinite article a:

- (4) a. Overt indefinite article (standard):
I told my employees, if I am talking to a customer don’t talk to me in Punjabi.
- b. Null indefinite article (nonstandard):
Then he thought, what about getting Ø girl [to marry] from India?

In the other set, the dependent variable was the absence of the definite article in contexts where Standard English requires the overt definite article the:

- (5) a. Overt definite article (standard):
Berkeley is just like a little India, because of the cosmopolitan area and the students.
- b. Null definite article (nonstandard):
Here I’m not working in Ø kitchen — yeah, I’m in Ø front desk; I’m a manager.

Most of the six independent internal (linguistic) variables coded for each article token derive from the hypotheses developed in the previous section. Other factors, such as grammatical function and modification, were included simply to include as wide a range of potential factors as possible:^v

Grammatical function (subject, VP object, PP object)
 Clausal topicality (theme, transition, rheme)
 Clause position (initial, medial, final)
 Specificity (specific, factive, non-specific, generic)
 Modification (bare, modified, quantified)
 Givenness (brand-new, new-anchored, unused, inferable, containing inferable, given)

Standard examples of each of these internal variables, taken from the interviews, are given in (6)–(11).

- (6) Grammatical Function
 - a. The future is safe if you get all the things. [subject]
 - b. My wife owns the store. [VP object]

- c. She's very involved in the community here. [PP object]

Three types of grammatical functions were distinguished; direct and indirect verbal objects were both coded under (6b). This factor is not discussed in detail as it did not appear to be a strongly conditioning factor; its relative effect is noted later in Table 8 and Table 9.

(7) Clausal Topicality

- a. If the economic situation would change I'm sure lot of Indian will go back. [theme]
- b. This is the major problem I am facing here. [transition]
- c. He had requested for a Hindi teacher. [rheme]

Following Young (1996), I coded three categories of clause topicality: the topic of the clause, the comment on the topic, and the linking or transitional term, if it occurred (Dik, 1978; Lambrecht, 1994; Birner and Ward, 1998). The transition tends not to be picked up as a new theme, once introduced, but rather serves a narrowing or specifying function. Degrees of discourse familiarity, as opposed to clausal topicality, were coded separately and discussed in (11).

(8) Clause Position

- a. The other brother was working in the factory. [initial]
- b. So that's the reason so many Indian family are separated. [medial]
- c. So six months I was there in the kitchen. [final]

Initial clause position included all NPs that occur at the left edge of the main clause. Final position, correspondingly, included NPs at the right edge of the clause, regardless of grammatical function. The category of medial position, also from Young (1996), was needed as NPs can appear in an intermediate object or subject position in a complex clause.

(9) Specificity

- a. There is a association that can help.^{vi} [specific]
- b. My Mom is a teacher. [factive]
- c. The parents will look for a girl. [non-specific]
- d. And you spend more time waiting for the bus. [generic]

Rather than using a binary or ternary distinction for specificity and genericity, I followed G. Sankoff and Mazzie (1991) in adopting the additional category of 'factive' for predicative NP constructions as in (9b), which perform neither a strictly specific function nor an entirely non-specific function, but rather mark category-membership. Sankoff and Mazzie treat factive as a variety of non-specific; however, in the present data factives patterned almost identically to specific in terms of null article use and was eventually grouped with specifics for the analysis. Most quantified NPs, such as a couple of people and a lot of Indians, were coded as non-specific.

(10) Modification

- a. I worked as a bartender. [bare]
- b. Bihar traditionally been a poor state. [modified]
- c. It's a problem for a lot of families. [quantified]

The category of 'quantified' in (10c) was fairly broadly interpreted, and included ranking adjectives (e.g. the first N, the last N, the best N, the whole N), quantifying phrases (e.g. a lot of N, a few of N), and numeral modifiers (e.g. the eleventh N). The basic property that these categories share is that their modifier either uniquely isolates the referent or quantifies the referent set. By contrast, the category of 'modified' included all other adjectivally modified NPs, in which the modifier simply restricts and specifies the potential real world referents.

(11) Discourse Givenness

- a. We have a Shiva temple. [brand-new]

- b. And it was a policy of the restaurant not to hire anybody with beard and long hair. [anchored-new]
- c. And then one day, the American government was calling dentist, doctor, nurses over here. [unused]
- d. They settle down over there so they have their own school. The professors were Indian and everything. [inferable]
- e. Because of the nature of the work, I'm always busy. [containing inferable]
- f. Some have a mentality of joining a start-up company... If the company goes to public, they become millionaires. [evoked]

For the final independent variable, I relied on Prince's (1981) Scale of Assumed Familiarity, further details of which are discussed later in relation to Hypothesis IV. NPs were classified into six degrees of discourse familiarity, exemplified in (11a-f), based on a scale of increasing familiarity status an NP may have at a given point in discourse. In (11a), the NP a Shiva temple is brand new in the discourse: it has not been mentioned previously and the hearer cannot anticipate it based on prior knowledge. In (11b) the new NP is anchored to a known discourse entity (the restaurant), and thus bears a slightly greater degree of familiarity to the hearer. In (11c) the American government is also new to the discourse but can be assumed by the speaker to be known to all participants in the discourse. In (11d) the professors is inferable information, as it has not been previously evoked but bears some prototypical relation to a familiar entity in the discourse and can therefore be anticipated by the hearer. (11e) contrasts with (11d) only in that the NP explicitly contains the evoked entity (the work) that provides the basis of the inference. Finally, in (11f), the second mention of the company is evoked: its referent has already been mentioned recently in the discourse and can be assumed to have a familiar status for both speaker and hearer.

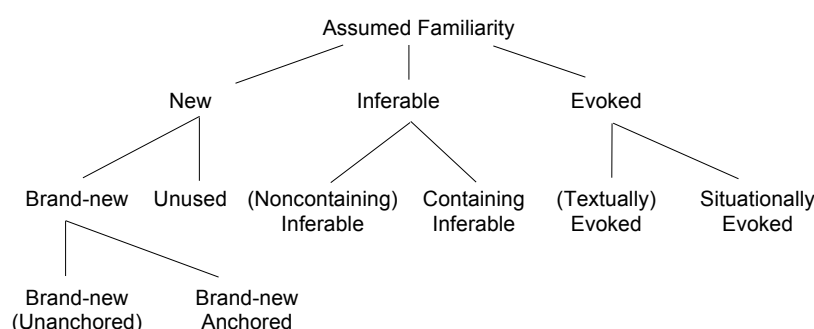


Figure 3: Taxonomy of Assumed Familiarity (Prince 1981: 237)

Based on an analysis of casual Standard English speech, Prince argues that the subdivisions in Figure 3 actually form a Scale of Assumed Familiarity, shown in (12):

- (12) Scale of Assumed Familiarity (Prince, 1981, p. 245):
 $E, E^S > U > I > I^C > BN^A > BN$

Prince uses this scale to account for discourse-driven choices of linguistic forms, such as articles, pronouns, subjects, or modifying phrases. For instance, if a discourse entity (e.g. John) is known to both speakers but is unused, then stating that John (U) bought a car is felicitous, whereas He (E) bought a car and A guy I know (I^C) bought a car are both infelicitous: he implies that the referent is accessible to the hearer, and a guy I know implies that the hearer does not know the individual in question.

The process of determining the givenness status of NPs is not always straightforward. Prince (1981, p. 244) acknowledges that the determination of an NP as evoked, unused, or inferable can sometimes be difficult. For instance, an NP such as my grandmother could be considered inferable on the basis that

individuals are known to have grandmothers, but it may also be considered unused information if the speaker and hearer both know the grandmother in question. To avoid variable coding, I only coded NPs as unused if they were culturally recognized entities such as the yellow pages, the internet or the U.S. government. For this reason, the category of unused has relatively few tokens.

Another problem between evoked and other, less familiar, states is the question of distance in time and discourse. The status of an NP as evoked ‘fades’ as the discourse progresses, until it can become unclear as to whether the NP is still evoked. Ariel (1990) uses a measure of distance to determine which parts of the discourse are still recent in the interlocutors’ minds; Lambrecht (1994) employs the notion of active (in the addressee’s memory) and accessible (available but distant in discourse) discourse referents. In general, it was possible to determine from context whether an NP was active at a given time.

In order to remove other ambiguities from Prince’s system, I was obliged to impose more explicit definitions on the categories of containing inferable and anchored new as well. I interpreted the category of containing inferable to include a number of NP structures: NPs with CP complements (e.g. the salesclerk that we hired), NPs with following locatives (e.g. the biggest market in this area), and ellipsis NPs (e.g. the first son was born last year and the second is...). The category of anchored new included NPs with an evoked locative (e.g. a man in the market), NP heads of relative clauses (e.g. a man I know), prepositionally modified NPs (e.g. a problem with the suppliers), and modified NPs relying on prior discourse knowledge (e.g. a newer store). Equative NP predicates, as in (9b), were classified as anchored as well, since the identity of the referent is being indexed with a discourse-familiar subject; the only exceptions to this were when the construction was used with irrealis or negative meaning or when the NP was predicated of an expletive subject. In such cases, the NP was classified as brand new.

The lexical semantics of particular verbs occasionally presented coding challenges. For instance, if a speaker uses the verb rent, the likelihood of the following NP falling within the class of typically rented items such as real estate or vehicles is high. Nevertheless, the NP cannot be treated as strictly inferable, as it is not a canonical or predictable extension of something already evoked in the discourse (as is the case with ‘bus... driver’). Only NPs that were very strongly conditioned by the verb were coded as inferable; the rest were coded as new.

Finally, a wide range of NPs were disregarded in both data sets based on ‘don’t-count’ criteria that I developed in the course of examining the data. As mentioned earlier, any contexts permitting optional article use in standard English were excluded. Proper names were excluded due to evidence of lexically specific article omission (e.g. the United States, the Niagara Falls). Articles within fixed constructions such as most of the N, all of the N, such a N, and in adverbial uses such as a little and a lot were also excluded, again due to the potential for lexically specific biases.

In order to evaluate each hypothesis across the continuum, the twelve individuals are arranged into three groups. Rather than dividing them evenly into three groups of four each, I followed the clustering of speakers that emerged in Table 1 and Table 2. Thus, among the three speakers in Group 1 (KD, SK, CK), English has the lowest functional and educational level in their repertoire; among the five speakers in Group 2 (RS, RR, KP, KK, GV) it is used regularly as a second language; and among the four speakers in Group 3 (RT, KB, SS, NT), English is often on an equal footing with other native languages. These groupings simply ease the presentation of data, and in the discussion that follows I focus on whether each hypothesis is confirmed for the group as a whole.

EVALUATION OF HYPOTHESES I-III

Evaluation of Hypothesis I: Transfer of L1 positional marking of discourse status

Hypothesis I anticipates that initial themes and final rhemes will be the categories favoring non-use of articles, as these are the unmarked contexts for old and new information respectively. Table 3 shows that no statistically significant pattern emerges for either of these predictions.^{vii} In fact, when separate chi-square tests are conducted for each group, only Group 1 has a significantly higher rate of article

omission with initial themes (χ^2 (df=1): 4.01, $p \leq 0.05$). This slight confirmation of the hypothesis is not repeated in Group 1's omission of articles with final rhemes, which in fact runs slightly counter to the prediction. Thus, although Table 3 does show that in terms of overall frequency rhemes tend to cluster in final position and themes in initial position, a link between these positions and article use is not evident.

Table 3. Null article use according to clause position and topicality

Group	Initial theme NPs		Noninitial theme NPs		Final rheme NPs		Nonfinal rheme NPs	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
1	38	74	4	25	47	53	29	59
2	101	50	33	48	333	44	87	44
3	68	29	22	36	180	16	85	15
Total	207	47	59	42	560	36	201	34

Note. Initial theme NPs versus noninitial theme NPs, $\chi^2(1) = 0.46$, $p = .4976$; final rheme NPs versus nonfinal rheme NPs, $\chi^2(1) = 0.23$, $p = .6315$.

Evaluation of Hypothesis II: Transfer of L1 article system

The first prediction of Hypothesis II is that the L2 English system will imitate the L1 system by only marking specific indefinites with an overt article. This is supported by the data in Table 4: we find a statistically significant difference across groups such that specific indefinite NPs, which would have an overt article in the first languages of the speakers, have rates of null marking that remain below 50% even for Group 1. On the other hand, non-specific indefinites, which are null marked in the first languages, are null marked at much higher rates.

Table 4. Null article use with definite and indefinite NPs

Group	Indefinite specific NPs		Indefinite nonspecific NPs		Definite NPs	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
1	29	48	42	81	83	46
2	129	31	206	57	343	38
3	116	16	132	20	209	18
Total	274	27	380	47	635	33

Note. Specific indefinite versus nonspecific indefinite NPs, $\chi^2(1) = 26.8$, $p < .001$.

However, the common claim that Indian speakers use one in place of a with specific indefinite NPs was not supported. Although individuals showed a noticeably higher rate of use of one in standard (emphatic, specific) reference than native English speakers might have, very few instances were clearly non-standard, most of which are given in (13).

- (13) a. Some entertainment workers are there. They are work in theatres for one dance.
 b. I went for a class. One Indian guy is there. Name is Anand.
 c. Like, I had one big joke, you know, on that part...
 d. There was one Britisher come over there.
 e. First of all, the Indians have one distinct accent to begin with.

Although Platt, Weber, and Ho (1984, p. 56) list this as an important transfer effect in their example in (2), the far more robust quantitative pattern in the present data is in fact the indirect transfer of function seen in Table 4, whereby a is restricted to a specificity marking function. The use of one may be more noticeable in Indian English, and thus cited more often in non-quantitative studies, due to its greater salience than its actual quantitative frequency.^{viii}

The second prediction of Hypothesis II, namely that the definite article will be absent because the L1 does not have one, is not clearly supported in Table 4. The prediction anticipates that at least Group 1 should have very high rates of non-use of definite articles; however, even for this group the rate of null definite articles does not exceed 50%.

Thus, we find Hypothesis II confirmed only in the part of the grammatical subsystem in which the first language has an overt form. Where a gap occurs in the first language (i.e. no definite article), we do not find a matching absence of use in the L2 English grammar; instead, there appears to be a near-even rate of overt and null use of the definite article.

Evaluation of Hypothesis III: Universal tripartite article system

Hypothesis III predicts a three-way distinction based on universal contrasts. Table 5 is slightly different from the previous tables as it gives rates for the predicted usage for each of these three categories: an overt form for [+HK, +SR]; an overt form for [-HK, +SR]; and a zero form for [-HK, -SR]. As Table 5 gives rates of predicted usage (rather than null usage, as in the previous tables), all rates are expected to be high, particularly for Group 1.

Table 5. Article use according to predicted universal tripartite system

Group	Overt <i>the</i> with definite NPs		Overt <i>a</i> with indefinite NPs		Null article with nonspecific/generic NPs	
	<i>n</i>	% overt	<i>n</i>	% overt	<i>n</i>	% null
1	72	56	29	52	53	75
2	248	60	129	69	301	50
3	193	81	116	84	148	19
Total	513	67	274	73	502	43

What we find, in fact, is that although Group 3 approaches the predicted high rates for the first two categories, since Standard English marks this distinction between [+HK] and [-HK], this is not the system that the less proficient groups universally adopt.

The only potential support for an emergent universal tripartite system is the high rate of null article use with non-specifics and generics that is apparent in Group 1. Group 2 and Group 3 decline in their rates of null use as they move towards the English system, which does not have null-marked non-specifics. However, even this single piece of support for the hypothesis can be equally well accounted for by Hypothesis II, as we have already seen that Group 1 speakers seem to be recreating the specific/non-specific distinction of their L1 in their L2 article use.

In sum, none of the first three hypotheses is completely confirmed. Hypothesis I showed a slight confirmation of null definite articles with clause-initial themes, but as this hypothesis treats discourse status in terms of salient contrasts in clause position, a partial confirmation of the hypothesis is insufficient. Hypothesis II was also partly supported within the domain of specificity marking, but not in terms of null use of definite articles. Hypothesis III only found support in the area of null marking of non-specifics, a pattern which can be just as well explained by Hypothesis II.

Taken together, the main finding thus far is that a transfer effect can be observed within the domain of specificity marking; variation in use of the definite article remains largely unexplained. An important implication of this difference is that transfer effects may be stronger in the part of the grammar where the first languages have an overt form, namely a specific indefinite article. In the next section, I explore the final hypothesis more closely, using a more refined model of discourse universals.

EVALUATION OF HYPOTHESIS IV: DISCOURSE FAMILIARITY

A fundamental problem with the approach to definiteness in the preceding discussion has been its treatment as a binary opposition. The near-even use of overt and null definite articles renders such binary models inadequate—a critique that has been made elsewhere (see Chesterman, 1991, p.39 for a summary). In this section, I use a more fine-grained model of givenness to consider the exploitation of discourse knowledge by speakers in their attempts to create order in their L2 grammars. G. Sankoff (1983, p. 245) cites the crucial role of agentive exploitation of linguistic contrast for discourse purposes by nonstandard users of a language:

It has been proposed that creolization involves people generating linguistic rules for which they have no evidence in the input. I suggest that what people do is reanalyze ‘grammatical’ input that is generated ‘discursively’—as the insertion of *pas* redundantly and emphatically in negative sentences in French was originally a discourse strategy that later became grammaticalized. The same sorts of strategies arise over and over in language, as people hit on the same solutions to their expressive problems, but rarely do these innovations become institutionalized—both first and second language learners must eventually conform to the preexisting norms of the languages they are learning. The genesis of both pidgins and creoles has taken place under conditions where the innovative strategies do get more of a chance to survive, because they are not competing in the same way with existing rules.

Indigenized non-native varieties share with creoles this widespread absence of native target norms, permitting a stabilization of discursive solutions.

Modelling givenness

Models of definiteness generally share the feature of ‘identifiability’ (Lambrecht, 1994; Lyons, 1999). Articles frequently begin their life as demonstratives or topic markers—highly discursive constructs—and may then gradually grammaticalize to cover a wider or narrower range of contexts (Greenberg, 1978; Givón, 1984; Lyons, 1999). For instance, while English has not extended the definite article to proper nouns or plural generics, Greek uses it with the former, and French with the latter (Trenkic, 2001, p. 109). It is possible that these discursive origins partly resurface through the process by which non-native speakers often ‘undo’ certain grammaticalizations and move towards transparent, analytic morphology.

Hypothesis IV explores whether there is a universal discourse effect governing article use in Indian English based on this function of identifiability. As identifiability can be thought of as scalar, the models of particular interest are those that distinguish among several types or degrees of definiteness or givenness, such as Givón (1984), J. Hawkins (1978) and Prince (1981); the suitability of these more fine-grained taxonomies for the study of non-standard article variation has been noted elsewhere as well (G. Sankoff and Mazzie, 1991; Y. Kachru, 2003).

Prince’s Scale of Assumed Familiarity, introduced earlier in relation to Figure 3, has a number of advantages for the present analysis: it focuses on discourse entities (or NPs) and so it can be applied directly to the question of article use; it appeals to universal discourse principles and thus represents, in an L2 situation, an alternative potential universal influence; and it makes implicational predictions for how discourse entities at different levels of the scale will relate to types of linguistic expressions, allowing us to move beyond binary contrasts. Hawkins’ (1978) taxonomy in some respects allows more detail than Prince’s (1981) model; for instance, his distinctions among associative anaphoric use (*a book... the author*), larger situation use (*in a village... the church*), and explanatory modifier use (*I remember the beginning of the war very well...*) are all subsumed under Prince’s category of inferables. However, his taxonomy does not order the various types that he identifies with respect to one another, a feature of particular interest to the present study.

Article use according to relative givenness

Standard examples of each type of discourse reference in Prince's model were given in (11) and non-standard null article examples from the present data set are given in (14):

- (14) a. Brand New (BN) We decided to rent \emptyset apartment.
 b. Anchored New (BN^A) They speak \emptyset different kind of English.
 c. Unused (U) You can browse through \emptyset internet.
 d. Inferable (I) He has one son and \emptyset second is on the way.
 e. Containing Inferable (I^C) From \emptyset first year of birth until...
 f. Evoked (E) In tenth the medium was English. After that I took science. \emptyset medium was English.

While Prince's scale underlies the Standard English article system to the extent that definite articles are associated with E, I^C, I, and U and indefinite articles are mainly associated with BN^A and BN, null realization of articles does not play a role in this variable usage.^{ix} Indian English appears to extend this functional differentiation to include null use as well. Rates of article omission relative to the discourse givenness of an NP are shown in Table 6. In order to temporarily factor out the first language effect on non-specific articles (see Table 4), Table 6 only includes definite and indefinite NPs with specific reference.

Table 6. Null article use according to familiarity status of NP

Group	BN		BN ^A		I ^C		I		E	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
1	11	36	15	53	13	46	50	40	11	73
2	29	7	87	34	62	35	126	37	57	60
3	42	5	73	22	74	8	81	17	23	43
Total	82	10	175	31	149	23	257	32	91	57

Note. $\chi^2(4) = 51.52, p < .001$.

As Figure 4 also shows, omission of the article, whether the or a, increases overall relative to the familiarity status of the NP. This statistically significant pattern holds within groups as well: Group 1 has predictably higher rates overall, followed by Group 2 and finally Group 3, the most proficient speakers. Brand new NPs have the lowest rates of article omission without exception, in each group, and evoked NPs have the highest rates of article omission, again without exception in each group. The intermediate categories of Prince's model—anchored new, inferable, and containing inferable—show intermediate and equivalent levels of article omission.

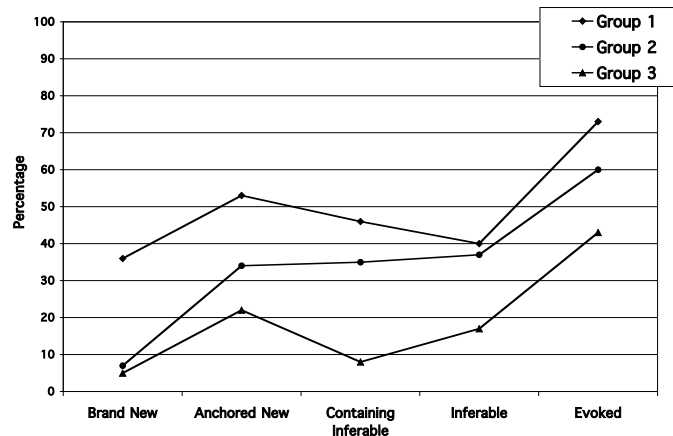


Figure 4: Null articles according to givenness (N values are given in Table 6)

Overall, Figure 4 indicates that a systematic use of overt articles mainly for purposes of discourse disambiguation may be operative for these Indian English speakers. The pattern shows that Prince's scale, which operates on other domains of Standard English syntax, can also extend to new grammatical domains in non-native discourse. It should be noted, however, that not all contrasts in Prince's scale appear to be equally salient to the speakers. The contrasts of new/anchored-new, new/inferable, and inferable/evoked show greater differences than inferable/containing-inferable or anchored-new/inferable. Also, the category of unused was highly variable, sometimes patterning like a new category and at other times resembling the rates for inferable NP types; as the number of unused NP tokens was low, it was excluded from Table 6.

Let us consider an example of this new pragmatic article use in detail. The extract in (15) is taken from a study of prosody in Indian English (Mishra, 1982) and although the study does not address article use at all, the actual speech extract used happens to include a rather striking demonstration of the alternations in marking of relative discourse givenness by Indian speakers. The fact that this extended illustration comes from the natural speech of an Indian speaker recorded for a completely different purpose lends support to the idea that this usage is widespread among Indian speakers.

(15) (from Mishra, 1982, p. 58)

When I had completed the training, ten day training at the language school, and you know that what happened there, there was another week for the vacation. And during that vacation I contacted the union and union person contacted his representative at the school. And that representative contacted the headmaster and headmaster had contacted the authority. But before that instance in the morning, first day of the term, I had met him and told him that I'm worried.

When I had completed <u>the training</u> ,	I
<u>Ø ten day training</u> at the language school	E
and you know that what happened there	
there was another week for <u>the vacation</u>	I
and during <u>that vacation</u>	E
I contacted <u>the union</u>	I
and <u>Ø union</u> person	I ^C
contacted <u>his representative</u> at the school	I ^C
and <u>that representative</u>	E
contacted <u>the headmaster</u>	I
and <u>Ø headmaster</u> had contacted the authority	E
but before that instance in <u>the morning</u> ,	I
<u>Ø first day</u> of the term, I had met him and told him that I'm worried.	E

This extract reflects the characteristic alternation of inferable and evoked NPs, a far more common pattern in natural dialogue than alternation between brand new and evoked, according to Prince. What is noteworthy here is the exceptionless alternation in article choice, whereby overt the or another definite determiner such as his are used with inferables (I and I^C), but either null the or that are used for subsequent evoked (E) references to the same entity. The shorter examples from my own data, given in (16), show a similar alternation such that a new or inferable referent includes the requisite article, following which an inferable or evoked reference to the previous referent omits the article.

- (16) a. I don't like the climate. Ø climate too much cold. [KD]
 b. It was a very small town I used to live. So you had to migrate to Ø other town for, like, after the seventh grade. [KP]
 c. Somehow he thought ki [that] if I'm marrying a girl, she should support him that I should continue enjoying my music. And then he thought, what about getting Ø girl from India? [KK]
 d. So in tenth, the medium English. After that I took science. Ø medium is English. [RT]
 e. They refer for the credit history. Where they don't have Ø credit history, there I'm not getting the credit card. So now I have to go for Ø secured credit card. [GV]

While these examples illustrate the importance of evoked status in article omission, they also point to other potential factors. For instance, the last NP in (16e) omits the article with the addition of the modifier secured, and (16a) and (16d) omit articles when the NP moves to a subject position. In order to evaluate whether it is in fact evoked status, more than any other factor, that is conditioning article omission, multivariate analysis is needed. Following a brief discussion of one remaining variable—NP modification—a summary of the relative influence of all factors is presented.

Further support for 'identifiability': NP modification and article omission

A final internal variable that showed a striking influence on article realization is type of modification in the NP. In fact, this pattern lends support to the preceding finding that NPs with unambiguous discourse reference can 'drop' their article more readily. If clarity of discourse reference and economy are major considerations for these speakers' decision to use or omit an article, then a modifier or a quantifier, which serves the function of specifying the referent from within a possible range, should render the article even more redundant and therefore more omissible. This is precisely the pattern found in the data.

Table 7. Null article use according to type of modification in NP

Group	Bare NPs		Modified NPs		Quantified NPs	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
1	83	43	45	58	26	92
2	379	29	186	49	113	78
3	250	16	124	17	83	28
Total	712	26	355	39	222	61

Note. $\chi^2(2) = 92.00, p < .001$.

Examples of quantified NPs without articles from the present data are given in (17); this influence of modifiers and quantifiers was observable in (15) as well.

- (17) a. I studied in Gujarati, but Ø second language was English. [RR]
 b. So it's more easier for new generation to understand Ø whole concept of the new technology and everything. [RR]
 c. And now he has one son and Ø second is on the way. [KK]
 d. In fact you find Ø lot of Andhra guys here rather than Bangalorians. [GV]
 e. So they went to India every year from day one. From Ø first year of birth till... [KB]

The statistically significant distribution in Table 7 and the selected examples in (17) indicate that while bare nouns require overt articles, modified nouns are more likely to be associated with omission of the article, and quantified nouns actually favor null articles. If modification restricts the potential real world referents of an NP token (e.g. the cold day), and quantification is understood as uniquely identifying the referents (e.g. the first day), then we can argue that exactly the same principle of disambiguation that was found for relative givenness lies behind this pattern as well.

In a quite different, more pedagogically-motivated analysis of article use among Indian speakers, Agnihotri, Khanna, and Mukherjee (1994, p. 185, 188) also find that articles are more frequently omitted when adjectives are present and when an NP includes a superlative quantifier. Though no clear explanation is given in their discussion, their results may reflect the pattern in Table 7. Thus, two of the most robust factors for omission of articles in the present data—NP modification and discourse givenness—both point to a single discourse pragmatic system of article use primarily for the purpose of disambiguation.

RELATIVE SIGNIFICANCE OF ALL FACTORS

The discussion in this paper has pointed to the particular importance of L1 transfer of specificity-marking, on the one hand, and of discourse ‘identifiability’, on the other, in determining article use among Indian English speakers. A multivariate analysis is presented in this closing section in order to further bolster the claim that these are the most influential of a range of factors. The statistical software package VARBRUL (Rousseau and D. Sankoff, 1978), used widely in sociolinguistics, has been introduced into the study of second language acquisition (Bayley and Preston, 1996; Berdan, 1996; Young, 1996) and has permitted a more careful consideration of classroom SLA, especially in cases of societal bilingualism (e.g. Mougeon and Beniak, 1991).

While the separation of speakers into three groups highlighted important proficiency-based differences earlier, for the VARBRUL analysis the entire group is treated as a single speech community. I separate null use of definite and indefinite articles in this analysis, as there may be instances in which a factor is relevant for one article but not the other, in which case combining the results for both articles would obscure the partial pattern. Table 8 presents the results for definite article use and Table 9 presents the results for indefinite article use. Since the stepwise regression takes into account the relative effect of all the other factors, the best indicator of the strength of a factor in each table is the column containing the VARBRUL weights. The relative magnitude of the weight relates to the relative strength of the effect of that factor on the dependent variable. A weighting below .5 indicates that the factor in question favors overt use of the article, while a weight greater than .5 means that the given factor favors omission of the article; a weight close to .5 means that the factor has little or no effect. Statistical significance or lack thereof, as determined by the regression analysis, is also indicated in this column through bracketing of values that were not found to be statistically significant by VARBRUL. The input value, or input probability, listed at the bottom of each table refers to the average probability that a rule will apply; thus, for instance, a weighting of more than .5 indicates that null article will occur with a greater frequency than 0.345. The independent factors in Table 8 and Table 9 are ordered according to their relative influence on the dependent variable.

In Table 8, the factors most strongly favoring null use of definite articles are discourse givenness and modification. Both of these factors also exhibit a scalar effect, where quantification favors null marking the most, modification slightly less so, and absence of all modification the least of all, and similarly evoked NPs strongly condition null marking, inferable NPs slightly less so, and brand new NPs least of all. The factors rejected by the multivariate analysis in Table 8, i.e. those that were not found to have a statistically significant impact on the use of definite articles, were grammatical function and specificity. The latter is not surprising, as specificity distinctions fall primarily within the domain of indefinite entities.

Table 9 shows that two of the three strongest factors in indefinite article are the same as for definite article use, namely discourse givenness and modification. The scalar pattern for these two variables is also repeated. However, additionally in Table 9 we see that specificity is a strong factor, reflecting the L1 transfer effect noted under Hypothesis II. The factors rejected as non-significant in Table 9 are grammatical function and clause position. The fact that clause position only appears to matter with definite articles also reflects the earlier finding under Hypothesis I that clause-initial themes favored null marking.

Table 8. VARBRUL results for definite article use (internal variables)

Variable	Tokens (n)	Null (n)	Null (%)	VARBRUL weight
Givenness				
Evoked	102	58	57	.779
Unused	54	18	33	.589
Inferable	323	98	30	.449
Cont. inferable	156	34	22	.371
Range				.408
Modification				
Quantified	100	51	51	.774
Modified	127	58	46	.654
Bare	408	99	24	.378
Range				.396
Clause position				
Initial	220	97	44	.683
Medial	146	33	23	.471
Final	269	78	29	.362
Range				.321
Clausal topicality				
Rheme	263	83	32	.597
Theme	210	87	41	.453
Transition	162	38	23	.403
Range				.194
Grammatical function				
Subject	203	89	44	[.582]
Verbal object	182	47	26	[.466]
Prep. object	250	72	29	[.457]
Range				.125
Specificity				
Generic	122	41	34	[.546]
Specific	513	167	33	[.489]
Range				.57

Note. Input value: .275; significance threshold: .05; values in brackets are nonsignificant.

Table 9. VARBRUL results for indefinite article use (internal variables)

Variable	Tokens (n)	Null (n)	Null (%)	VARBRUL weight
Specificity				
Generic	14	13	93	.971
Nonspecific	366	164	45	.564
Factive	183	56	31	.397
Specific	91	17	19	.325
Range				.646
Modification				
Quantified	122	84	69	.818
Modified	228	80	35	.489
Bare	304	86	28	.361
Range				.457
Givenness				
Inferable	68	46	68	.804
Anchored-new	313	135	43	.544
Brand-new	273	69	25	.365
Range				.439
Clausal topicality				
Theme	56	36	64	.735
Rheme	498	185	37	.502
Transition	100	29	29	.354
Range				.381
Grammatical function				
Subject	29	20	69	[.633]
Verbal object	504	191	38	[.506]
Prep. object	121	39	32	[.443]
Range				.190
Clause position				
Final	400	156	39	[.527]
Medial	213	70	33	[.472]
Initial	41	24	59%	[.383]
Range				.144

Note. Input value: .345; significance threshold: .05; values in brackets are nonsignificant.

The factors that influence use of definite and indefinite articles thus form a relatively coherent set, and the statistical regression shows that in fact the strongest factors are indeed givenness and specificity, outweighing such factors as clause position and grammatical function. This type of statistical analysis can thus settle issues that have remained unclear in previous studies; for instance, the relative importance of topichood as opposed to subjecthood when high rates of clause-initial null articles are apparent (Huebner, 1983; Parrish, 1987).

DISCUSSION AND CONCLUSION

This study has aimed to advance our understanding of new non-native dialects of English through a close analysis of factors in grammatical divergence. The initial implicational scaling found that certain features diverge much more dramatically than others, and the markedly different article system seems to be sensitive to specificity but also to the relative discourse familiarity of the NP. These findings promise a more complete account of why Indian speakers vary in their use of articles, as the study has isolated not only the more commonly cited L1 effects, but also new pragmatic functions.

Klein (1980, p. 77), discussing variation in the use of synthetic and analytic forms of progressive aspect among Spanish-English bilinguals, observes that ‘the actual choice between the more precise and the less precise alternatives... should be determined by pragmatic strategies based on relative need for precision.’ Klein and Perdue (1992, p. 311) also argue that individuals with different L1s have a ‘basic variety’ at one stage of learning ‘which seems to represent a natural equilibrium between the various phrasal, semantic, and pragmatic constraints’. This integral role of pragmatics also echoes Tarone and Parrish’s (1988) finding that articles with referential definites were employed with greater native-like accuracy in narrative genres due to the greater communicative burden of precise and efficient reference.

The capacity for more given entities to drop their articles has been noted in a number of other studies, with varying degrees of explanation (G. Sankoff and Mazzie, 1991; Agnihotri, Khanna, and Mukherjee, 1994; Bruyn, 1995; Jarvis, 2002). Jarvis (2002, p. 416) explicitly finds this constraint to be active: ‘In the Finns’ data, on the other hand, the use of Ø probably does not represent a simplified register as much as it represents the L1 Finnish convention of avoiding (what Finns perceive to be) redundant markers of definiteness and indefiniteness when these properties of an NP are already salient in a given discourse context.’ He cites Givón’s description of the relatively greater burden on processing of newer information, a factor that appears to be strongly reflected in the present data as well: ‘more continuous, predictable, non-disruptive topics will be marked by less marking material; while less continuous, unpredictable/surprising, or disruptive topics will be marked by more marking material’ (Givón, 1984, p. 126).

Studies of creoles have also found null marking with definites, despite the ‘bioprogram’ prediction that null-marking will be associated with generics. Bruyn (1995, p. 73) notes in passing that, counter to Bickerton’s predictions, null marked NPs in Sranan are not always generic, and can also be definite. G. Sankoff and Mazzie (1991, p. 7) also find definite NPs that are null marked in Tok Pisin, and they cite Corne (1977, p. 14) who, for Seychelles Creole, also noted that definite nouns may be unmarked ‘when no ambiguity is possible’.^x Sankoff and Mazzie (1991) furthermore observe that both Bickerton and Givón themselves have instances of null-marked definites in their own data but explicitly choose to exclude them. Givón discounts these instances as referring to characters peripheral to the main story; Bickerton, using somewhat circular reasoning, chooses to discard such cases because they do not fit his definition: ‘A total of 260 zero articles are found in the texts under consideration. 24 of these—almost one-tenth—may immediately be discarded, since they refer to things which have definite reference.’ (Bickerton, 1977, p. 235). This is not to say that null-marking cannot be associated with generics, but given the present data and the considerable range of studies that have made passing reference to the occurrence of null article use with given NPs, the availability of a competing discourse principle for the omission of articles cannot be ignored.

As Jarvis (2002, p. 414) notes, however, the marking of broad discourse distinctions such as these in second language English is often triggered, nevertheless, by the existence of differences between the L1 and L2 systems. Thus, although discourse knowledge plays a role in the restructuring of these Indian speakers' English grammars, these forces may only intervene due to the initial disruption of significant mismatches between their L1s and English and the resulting need to settle on some principled basis for using article forms.

To conclude, this study has shown, first, that systematic divergence in a stable non-native variety can indeed be identified in quantitative terms. Furthermore, the results indicate that rather than acting as opposing forces, language transfer and universals may enter into complementary partnerships. In the present case, the existence of an overt specific article in individuals' L1s appears to have a strong influence on their use of the English indefinite article, but their L1 lack of a definite article invites the imposition of other discursively available pragmatic principles for the use of English articles. As a result, the influence of factors that seemed, on the surface, to be unrelated, namely discourse familiarity and NP modification, can also be subsumed under a single pragmatic explanation.

ACKNOWLEDGEMENT:

I am indebted to John Rickford, Elizabeth Traugott, Arnold Zwicky, Ishtla Singh, and the SSLA editors and anonymous reviewers for much helpful input at earlier stages of this research. I also received valuable comments from audiences at the LSA, Atlanta (January 2003) and at UC Davis and UC San Diego (February 2003).

REFERENCES

- Agnihotri, R. K., Khanna, A. L. & Mukherjee, A. (1994). The use of articles in Indian English: Errors and pedagogical implications. In R. K. Agnihotri & A. L. Khanna (eds.). *Second language acquisition: Sociocultural and linguistic aspects of English in India* (pp. 178-198). New Delhi: Sage.
- Andersen, R. (1978). An implicational model for second language research. *Language Learning*, 28, 221-282.
- Ariel, M. (1990). *Accessing NP antecedents*, London: Routledge.
- Bamgbose, A. (1998). Torn between the norms: innovations in world Englishes. *World Englishes*, 17/1, 1-14.
- Bayley, R. (1994). Interlanguage variation and the quantitative paradigm: Past tense marking in Chinese-English. In E. Tarone, S. Gass & A. Cohen (eds.). *Research methodology in second language acquisition* (pp. 157-181). Hillsdale, New Jersey: Erlbaum.
- Bayley, R. (1999). The primacy of aspect hypothesis revisited: Evidence from language shift. *Southwest Journal of Linguistics*, 18.2, 1-22.
- Bayley, R. & Preston, D. (eds.). (1996). *Second language acquisition and linguistic variation*. Amsterdam: John Benjamins.
- Berdan, R. (1996). Disentangling language acquisition from language variation. In R. Bayley & D. Preston (eds.) *Second language acquisition and linguistic variation*. (pp. 203-245). Amsterdam: John Benjamins.
- Bhatia, T. (1993). *Punjabi: A cognitive-descriptive grammar*. London: Routledge.
- Bickerton, D. (1971). Inherent variability and variable rules. *Foundations of language*, 7, 457-492.
- Bickerton, D. (1977). Pidginization and creolization: Language acquisition and language universals. In A. Valdman (ed.), *Pidgin and creole linguistics* (pp. 49-69). Indiana University Press.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma.
- Birner, B., & Gregory W. (1998). *Information status and noncanonical word order in English*. Amsterdam: John Benjamins.
- Bruyn, A. (1995). *Grammaticalization in creoles: the development of determiners and relative clauses in Sranan*. Amsterdam: IFOTT

- Cardona, G. (1965). *A Gujarati Reference Grammar*. Philadelphia: University of Pennsylvania Press.
- Chaudron, C., & Parker, K. (1990). Discourse markedness and structural markedness: The acquisition of English noun phrases. *Studies in Second Language Acquisition*, 12, 43-63.
- Cheshire, J. (ed.). (1991). *English around the world*. Cambridge: Cambridge University Press.
- Chesterman, A. (1991). *On Definiteness: A Study with Special Reference to English and Finnish*. Cambridge: Cambridge University Press.
- Corne, C. (1977). *Seychelles creole grammar: Elements for Indian Ocean proto-creole reconstruction*. Tuebingen: TBL-Verlag Narr.
- Dik, S. (1978). *Functional grammar*. Amsterdam: North-Holland Press.
- DeCamp, D. (1971 [1968]). Toward a generative analysis of a post-creole speech continuum. In D. Hymes (ed.). *Pidginization and creolization of languages* (pp. 349-370). Cambridge: Cambridge University Press.
- Dulay, H. C. & Burt, M. K. (1974). Errors and strategies in child second language acquisition. *TESOL Quarterly*, 8, 129-136.
- Fodor J. D. & Sag, I. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy*, 5, 355-398.
- Givón, T. (1984). The pragmatics of referentiality. In D. Schiffrin (ed.). *Georgetown University round table on language and linguistics* (pp. 120-138). Washington, DC: Georgetown University Press.
- Goto Butler, Y. (2002). Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition*, 24, 451-480.
- Greenberg, J. (1978). *Universals of human language*. Stanford: Stanford University Press.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English. *Journal of Linguistics* 3, 37-81.
- Hawkins, J. (1978). *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm.
- Hawkins, R. (2001). *Second Language Syntax: A Generative Perspective*. Oxford: Blackwell Publishers.
- Ho, M. L. & Platt, J. (1993). *Dynamics of a contact continuum: Singaporean English*. Oxford: Oxford University Press.
- Hudson, T. (1993). Nothing does not equal zero: Problems with applying developmental sequence findings to assessment and pedagogy. *Studies in Second Language Acquisition*, 15, 461-493.
- Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Ann Arbor, MI: Karoma.
- Ionin, T., Ko, H. & Wexler, K. (2003). Specificity as a grammatical notion: Evidence from L2-English article use. *WCCFL 22 Proceedings* (pp. 245-258). Somerville, MA: Cascadilla Press.
- Jarvis, S. (2002). Topic continuity in L2 English article use. *Studies in Second Language Acquisition*, 24, 387-418.
- Kachru, B. (1965). The indianness of Indian English. *Word*, 21, 391-410.
- Kachru, B. (1983). *The indianization of English: The English language in India*. Oxford: Oxford University Press.
- Kachru, Y. (2003). On definite reference in world Englishes. *World Englishes* 22/4, 497-510.
- Kandiah, T. (1991). South Asia. In J. Cheshire (ed.). *English around the world* (pp. 271-288). Cambridge: Cambridge University Press.
- Klein, F. (1980). A quantitative study of syntactic and pragmatic indications of change in the Spanish of bilinguals in the U.S. In W. Labov (ed.), *Locating language in time and space* (pp. 69-82). New York: Academic Press.
- Klein, W. & Perdue, C. (1992). *Utterance structure: developing grammars again*. Amsterdam: John Benjamins.
- Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: Cambridge University Press.
- Lowenberg, P. (1986). Non-native varieties of English: Nativization, norms, and implications. *Studies in Second Language Acquisition*, 8, 1-18.
- Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- McGregor, R. S. (1995). *Outline of Hindi Grammar with exercises*. Oxford: Oxford University Press.
- Mishra, A. (1982). Discovering connections. In J. Gumperz (ed.) *Language and social identity* (pp. 57-72). Cambridge: Cambridge University Press.
- Mesthrie, R. (1992). *English in language shift*. Cambridge: Cambridge University Press.

- Mohan, T. (1994). *Argument structure in Hindi*. Stanford: CSLI Publications.
- Mougeon, R., & Beniak, É. (1996). Social class and language variation in bilingual speech communities. In D. Schiffrin, G. Guy, J. Baugh, & C. Feagin (eds.), *Towards a Social Science of Language: a Festschrift for William Labov*, vol. 1 (pp. 69-99). Amsterdam: John Benjamins.
- Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning*, 37, 361-383.
- Pienemann, M., Johnston, M. & Meisel, J. (1993). The multidimensional model, linguistic profiling, and related issues. *Studies in Second Language Acquisition*, 15, 495-503.
- Platt, J., Weber, H. & Ho M. L. (1984). *The New Englishes*. London: Routledge.
- Politzer, R. (1976). An empirical study of children's ESL development and rapid profile. In P. McKay (ed.), *ESL development: Language and literacy in schools* (pp. 115-259), Vol. 2. Amsterdam: John Benjamins.
- Prince, E. (1981). Toward a taxonomy of given-new information. In Peter Cole (ed.) *Radical pragmatics* (pp. 223-255). New York: Academic Press.
- Rickford, J. (2002). Implicational scales. in J. Chambers, P. Trudgill & N. Schilling-Estes (eds). *Handbook of language variation and change* (pp. 142-167). New York: Blackwell Press.
- Rousseau, P., & Sankoff, D. (1978). Advances in variable rule methodology. In D. Sankoff (Ed.), *Linguistic variation: models and methods* (pp. 57-69). New York: Academic Press, Inc.
- Sahgal, A. (1991). Patterns of language use in a bilingual setting in India. In Cheshire, J. (ed.). *English Around the World* (pp. 299-308). Cambridge: Cambridge University Press.
- Sankoff, G. (1983). Comments on Valdman's 'Creolization and second language acquisition'. In R. Andersen (ed). *Pidginization and creolization as second language acquisition* (pp. 235-240). Rowley: Newbury House Publishers.
- Sankoff, G. & Mazzie, C. (1991). Determining noun phrases in Tok Pisin. *Journal of Pidgin and Creole Languages*, 6, 1-24.
- Schiffman, H. (1999). *A Reference Grammar of Spoken Tamil*. Cambridge: Cambridge University Press.
- Sharma, D. (2005). Dialect stabilization and speaker awareness in non-native varieties of English. *Journal of Sociolinguistics*, 9:2, 195-225.
- Sridhar, K. K. (1985). Sociolinguistics of non-native English. *Lingua*, 68, 39-58.
- Sridhar, S. N. (1990). *Kannada*. London: Routledge.
- Tarone, E. & Parrish, B. (1988). Task-related variation in interlanguage: The case of articles. *Language Learning*, 38, 21-45.
- Trenkic, D. (2001). Establishing the definiteness status of referents in dialogue in languages with and without articles. *Working papers in English and applied linguistics*, University of Cambridge.
- Vallduv, E. (1992). *The informational component*. Garland, New York.
- Williams, J. (1987). Non-native varieties of English: A special case of language acquisition. *English World-Wide*, 8, 161-199.
- Wolfram, W. (1985). Variability in tense marking: A case for the obvious. *Language Learning*, 35, 229-253.
- Young, R. (1996). Form-function relations in articles in English interlanguage. In R. Bayley & D. Preston (eds), *Second language acquisition and linguistic variation*, (pp. 135-175). Amsterdam: John Benjamins.

ENDNOTES:

ⁱ Only the rates of null usage for evoked definite articles are listed in Table 2, i.e. the use of definite articles with NPs that have already been mentioned in the discourse. As discussed later in this study, Prince (1981) treats this as the most ‘given’ status for an NP and it is this category that exhibits the highest rates of article omission.

ⁱⁱ A comparison with Table 1 shows that the external (social) factors correlating most closely with these rates of nonstandardness are educational background and functional use of English, thus approximating a proficiency-based continuum (see Sharma 2005).

ⁱⁱⁱ All coding was done by the author. In order to check for reliability, a random selection of 100 NPs were coded independently by a trained researcher. Before discussion between coders, the average inter-rater reliability was 97.9% for standardness judgements on copula, agreement, past tense, and articles, and 93.25% for the six internal factors coded for article tokens.

^{iv} For the purposes of this study, Standard British and American English are treated as generally equivalent systems. This is not always the case, as in dialectal variants such as in (the) hospital. Such cases would have been omitted from the data set, but in fact none were encountered during coding.

^v External (social) factors were also coded but are discussed separately in Sharma (2005).

^{vi} Variation in a and an occurs in the data but is not included as a variable. a is occasionally generalized to positions that standardly require an.

^{vii} The total number of tokens in Table 4, Table 5, and Table 7 is 1289. This total is slightly lower in Table 3 and Table 6. This is because in Table 3 the category of ‘transition’ is omitted (only ‘theme’ and ‘rheme’ are relevant), and in Table 6 only specific NPs are counted and the category of ‘unused’ is omitted due to low N values.

^{viii} Another adaptation of an English form to a substrate function occurs specifically in the speech of the two South Indian speakers. Kannada and Tamil both allow demonstrative and quantifier forms to be used as indefinite articles, and the two speakers of these languages frequently used some and this as specific indefinite articles, and that as a definite article.

^{ix} Discourse-based article omission is not completely absent in native varieties, but is highly restricted to certain types of subjects: definite subject articles (Ø Last film I watched was Star Wars) and indefinite subject articles (Ø Man walks into a bar...) are occasionally omitted. Articles with objects are not omissible in Standard English casual speech, whether indefinite (*A man wants Ø gift for Ø 3-year-old kid) or definite (*I told Ø man that the shop was closed).

^x Sankoff and Mazzie (1991) also employ Prince’s model in order to access the role of discourse status distinctions in article choice in Tok Pisin. However, their study is not directly comparable to the present one, as their focus is on the specialization of different forms (e.g. dispela, ia, wanpela) for different functions, mirroring a pattern shared by the Austronesian substrate languages.